졸업논문
소개

음성·얼굴 움직임 데이터의 상관 구조 분석 기반의 학습 방법론에 대한 연구

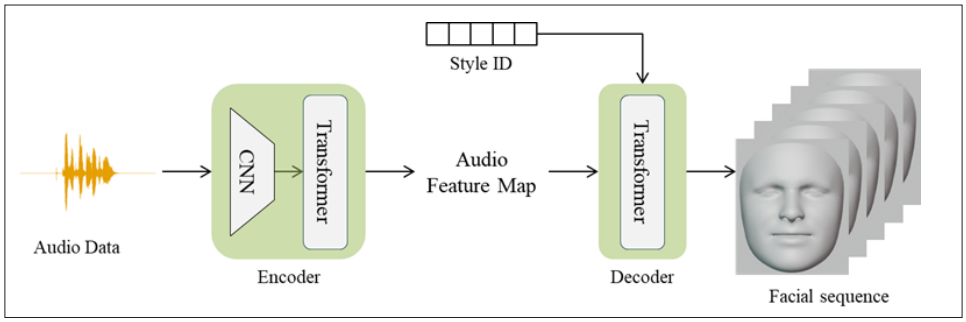
임성원 / 광운대학교 Intelligent Computing Lab.

음성만을 입력으로 가상 인간의 얼굴 움직임을 생성하는 기술인 음성 구동 얼굴 애니메이션(Audio-Driven Facial Animation)은 별도의 실사 모션 캡처 없이도 자연스러운 입술 움직임과 세밀한 감정 표현을 구현할 수 있어 디지털 휴먼 제작 분야에서 큰 주목을 받고 있다. 본 논문은 음성-얼굴 움직임의 관계성을 정량평가하고 이를 기반으로 음성과 개별 표정 기저 가중치 간 상호 연관성을 반영한 음성과 표정 정보를 학습하는 딥러닝 모델의 파이프라인을 제안한다.

음성 구동 얼굴 애니메이션이 표현하는 얼굴 모션 캡처 데이터는 대표적으로 시간에 따라 변화하는 얼굴 메시(Mesh) 또는 2D·3D 랜드마크의 좌표를 직접 기록하는 기하(Geometry) 기반 표현과 이러한 기하 정보를 여러 개의 기준 표정으로 압축하여 선형 결합 계수로 다루는 블렌드셰이프(Blendshape) 기반 계수 시퀀스 형태로 표현되며, 시간에 따라 변화하는 좌표와 계수를 통해 얼굴 각 부위의 근육 움직임을 정밀하게 기술할 수 있다. 음성 구동 얼굴 애니메이션은 별도의 실사 모션 캡처 장비 없이도 입술 움직임과 표정 변화를 추가적인 공정 없이 합성할 수 있으며, 딥러닝 학습 방법의 발전과 함께 〈그림 1〉과 같은 형태로 서로 다른 형태와 특성을 가진 데이터 간의 관계성을 학습하는 크로스모달(Cross-Modal) 학습 방식이 도입되면서, 단순히 발음에 따른 발성 시 입 움직임을 모방하는 수준을 넘어, 화자별 억양 차이에 따른 미세한 입 움직임의 변화와 감정에 따른 전체적인 얼굴 표정 변화를 동시에 반영하는 현실적인 사람 얼굴 표정을 재현하려는 연구가 활발히 진행되고 있다. 그러나 이러한 기술적 진전에도 불구하고, 음성 구동 얼굴 애니메이션 연구에서 음성 신호와 캡처된 얼굴 모션 사이의 대응 구조, 그리고 그 관계가 개별 얼굴 근육 운동 또는 블렌드셰이프 성분 단위로 어떻게 표현·분해되는지에 대한 정량적 분석은 여전히 부족하다.

본 연구는 한계점을 극복하기 위해 음성 특징과 얼굴 근육 움직임을 정량화하여 분석한다. 첫 번째로 동일 내용을 발화한 블렌드셰이프 모션 데이터에 대한 동일 계수 간의 표현 유사도를 ICC(3,1)를 통해 동일 발화에 대해 개별 근육 움직임 데이터가 얼마나 일관적인 표현을 하는지 평가한다. 두 번째로 음성과 개별 블렌드셰이프의 개별 선형 관계를 정량화하여 근육 움직임 데이터의 일관적인 표현과 음성 데이터 간의 연관성을 피어슨 상관계수, 정준

졸업논문 소개



<그림 1> 음성·스타일 조건 3D 얼굴 애니메이션 인코더-디코더 구조

상관분석(Canonical Correlation Analysis, CCA)을 통해 분석한다. 이를 통해 음성 구동 얼굴 애니메이션 생성 문제가 음성에 따라 서로 다른 통계적 특성과 의미적 역할을 지니는 여러 블렌드셰이프 계수를 동시에 예측해야 하는 멀티 태스크(Multi-Task) 문제임을 식별한다. 그럼에도 기존 연구 대부분이 단일 손실 함수를 사용해 모든 계수를 동일하게 최적화함으로써 각 근육 움직임을 학습하는 태스크들이 서로의 학습을 방해하는 부정 전이(Negative Transfer)가 발생하고 있음을 태스크 간의 학습 기울기와 크기 차이의 정량화를 통해 보여준다.

이러한 분석을 바탕으로, 음성과 개별 표정 기저 가중치 간 상호 연관성을 반영한 음성과 표정 정보를 단일 파이프라인으로 학습하는 구조를 설계한다. 제안 모델은 음성과 관련성이 높은 블렌드셰이프에 더 높은 가중치를 부여하고, 관련성이 현저히 낮은 블렌드셰이프를 정량적 평가를 통해 배제함으로써 계수 간 부정 전이를 완화함으로써 안정적이고 효율적으로 학습하도록 유도한다.

그 결과를 직접 반영한 ICC 기반 태스크 선택·가중 전략이 구강·턱 계열 계수의 예측 성능을 향상시킨다는 점을 보였다. 이와 같은 "분석 → 갈등 구조 파악 → 손실 설계로의 반영" 절차는, 향후 음성 구동 3D 얼굴 애니메이션 뿐 아니라 유사한 크로스모달 다중 과제 문제에서 태스크별 특성을 고려한 학습 전략을 설계하는데 참고 가능한 분석적 틀을 제공한다.

임 성 원

- 2024년 : 광운대학교 전자재료공학과 학사
- 2026년 2월 : 광운대학교 전자재료공학과 석사졸업(예정)
- 주관심분야 : 디지털 휴먼, 생성모델, 컴퓨터 비전